

A BILINGUAL ACADEMIC WORD LIST: THE MERGING OF A NORWEGIAN AND A SWEDISH LIST

SOFIE JOHANSSON, KRISTIN HAGEN AND
JANNE BONDI JOHANNESSEN

ABSTRACT

This paper describes the work and methods used to compile monolingual Norwegian and Swedish academic word lists as well as a merged Norwegian-Swedish list. The resulting list is discussed with respect to similarities and especially differences between the two languages, in terms of concepts such as cognates, false friends and remote friends.

[1] INTRODUCTION

Recently, two new Scandinavian academic vocabulary lists have been developed; one for Swedish (Jansson et al. 2012) and one for Norwegian (Johannessen et al. 2016). There is a definite need for such lists, even beyond the two individual lists, and since the Scandinavian languages are very similar linguistically, and the attitudes to the linguistic situation are also the same (see Sections 1.1–1.3), the present merged Swedish-Norwegian list will be a fundamental tool for higher education and other institutions.

Both individual lists were compiled using methods commonly used in language technology and corpus linguistics. In this paper, we present the merging of these two lists into a common list of academic vocabulary. After a brief description of previous work on English vocabulary lists in Section 2, the work and methods used for developing the two monolingual Scandinavian lists are presented in Section 3. The two lists display some differences that are probably due to the compilation methods applied to produce the lists, and this section is intended to help the reader better understand these differences.

In spite of the similarities between the two languages both socially and linguistically, there are some differences between them. In Section 4, we illustrate many aspects of the two languages by comparing the merged list. While some words are naturally the same, others are very different, and do not even exist in the other language, and yet others are cognates to a varying degree. One

problem for those who know only one of the languages is the large number of false friends, and even, as we call them, remote friends. The paper is concluded in Section 5.

Academic language in this paper is understood as that which especially occurs in academic publications, i.e. journal articles, PhD theses, MA theses, research reports, and university level textbooks.

[1.1] *Swedish and Norwegian*

Swedish and Norwegian are two closely related languages in the North Germanic branch of the Germanic language family. Most dialects are mutually intelligible not only within Sweden (and the Swedish-speaking part of Finland) and Norway, but also across the borders of these countries.

A short excerpt from the UN General declaration of human rights illustrates the similarity between the two languages.

Alla människor äro födda fria och lika i värde och rättigheter. De äro utrustade med förnuft och samvete och böra handla gentemot varandra i en anda av broderskap. (Swedish)

Alle mennesker er født frie og med samme menneskeverd og menneskerettigheter. De er utstyrt med fornuft og samvittighet og bør handle mot hverandre i brorskapets ånd. (Norwegian)

All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood. (English)

However, since the languages are spoken in separate countries, they have developed their own characteristics, not just with respect to grammar, but also the lexicon. Since the countries have their own institutions at all levels of society it is expected that the academic vocabulary will be somewhat different in the two countries. In order for students from one of the languages to be able to read books in the other language, or even study in the other country, an academic word list will be useful. As far as we know, monolingual academic word lists exist for various languages such as English, Norwegian and Swedish, and word lists translated from an English academic word list into a target language can also be found, but there are no bilingual word lists compiled by two separately constructed monolingual word lists. See appendix 1 for further references to academic word lists.

[1.2] *Language policies in Scandinavian higher education*

Language legislation and policies in the Scandinavian countries state that the Scandinavian languages have equal status. A language Act specifies language policies in authorities and public organizations, and the universities interpret these national language Acts in similar ways. This means that teaching and textbooks can be in Swedish as well as in Norwegian, and vice versa. When the academic language is as different as we show it to be in this paper, it is clear that an easily accessible common, bilingual wordlist will be highly useful.

[1.3] *Why a common, bilingual academic word list?*

The question "Why a common, bilingual academic word list" really consists of two questions: Why a bilingual list? and Why a common one? We will answer these in turn. The need for a bilingual list is based on the factors given in Sections 1.1 and 1.2. While the two languages are very similar, there are differences that have arisen due to the fact that both languages serve as national languages in their respective countries. There are therefore pairs of words, equivalents in both languages, whose members are so different from each other that those not knowing the language well would have problems (Section 4.1). In addition, there are so-called false friends and what we call remote friends (Section 4.2), which could easily fool readers who are not aware of the differences. Given that the Scandinavian countries are small, it is not uncommon for academic textbooks to be available in only one of the Scandinavian languages. A bilingual academic word list will therefore be an important tool for students in these countries. The existence of such lists will also be an argument for teaching institutions to avoid English textbooks, and therefore facilitate students' abilities to study in their mother tongue.

The need for compiling a common list is due to the fact that the two previous monolingual lists did not overlap completely. Each of those lists comprises 750 words, while the common, compiled list covers 975 words.

[2] BACKGROUND AND PREVIOUS RESEARCH

This section provides a historical introduction to academic word lists as well as a summary of existing academic word lists.

[2.1] *English academic word lists*

In 1921, the American psychologist Edward Thorndike needed an English word list for educational purposes. He therefore made a word list of the 10,000 most common words in general English. Another commonly used word list for English is the General Service List (GSL) compiled by West in 1953. GSL is a set of

2,000 words selected to be of the greatest "general service" to learners of English. The University Word List (UWL) was constructed by Xue and Nation in 1984. It is a list of 836 words that are not included in the 2,000 words of the GSL, but are common in academic texts. The UWL accounts for 8% of the words in a typical academic text according to Nation (1990, p. 19). The Academic Word List (AWL), developed by Coxhead (2000), is perhaps the academic word list most frequently referred to. The AWL contains 570 word families which were selected according to certain principles of frequency and distribution and does not include words belonging to the most frequent 2,000 words of English. The Academic Vocabulary List (AVL), developed by Gardner & Davies (2013), consists of lemmas instead of word families. The authors mention several advantages of using lemmas in a word list used for pedagogical purposes. One of the most important ones is the need that less proficient speakers (second language learners and children of school age) have for separating derivational forms of lemmas when learning new words. Gardner & Davies refer to similar claims by Nippold and Sun (2008: 365), Nagy (2007) and Schmitt and Zimmerman (2002). Apart from the wordlists already mentioned, see Appendix 1 for a summary of some more existing academic word lists.

[3] METHODS FOR DEVELOPING ACADEMIC WORD LISTS AND FOR MERGING LISTS

This section summarizes the general principles of compilation methods used to produce academic vocabulary lists. The compilations of the two individual monolingual Scandinavian academic vocabulary lists are described and the merging of the two lists is explained in detail.

[3.1] *Technical terms*

In order to explain the processes, methods and procedures to describe the compilation and merging of academic word lists, a brief summary of commonly used technical terms is provided in appendix 2.

[3.2] *Selection principles for monolingual academic word lists*

In order to find the academic words of a language, a corpus of academic texts can be used. There are three main processes that must be used to identify academic words from an academic corpus (cf. Gardner & Davies 2013, Coxhead 2000):

- (i) Exclusion of high-frequency non-academic words that can be found in every corpus

- (ii) Exclusion of subject-specific terminology that has a high frequency in only some parts of the academic corpus
- (iii) Inclusion of words common in several subjects

[3.3] *The Swedish academic word list*

The Swedish academic word list consists of 652 words and is extracted from a corpus of more than 25 million words. The corpus is compiled of theses and articles from academic fields within humanities and social sciences from the database SwePub. The fields in humanities are represented by texts from ethnology, philosophy, history, arts, literary science, religion and linguistics. The fields in social science include texts from economics/industry and commerce, law, media and science of communication, psychology, social and economic geography, sociology, political science, media and educational science (Jansson et al., 2012, Sköldbberg & Johansson Kokkinakis, 2012, Ribbeck et al., 2014).

The method used in compiling the Swedish academic word list is based on frequency and frequency distribution. Reduced frequency is used to account for a word's frequency in a corpus divided into subsections. The sections are not related to genre or subject-specific text. Further, keyness (Scott & Tribble, 2006) was used to select words which were representative to the texts compared to other texts, e.g. novels, which were used as reference texts. The keyness score was set to 1.1. Another criterion used was range, which would ensure that the words were common in all subjects and not only one. The requirement was that the word should appear at least 15 times per million tokens in each of the university subjects. The third selection principle was to use a corpus of easy-to-read texts (LäsBarT) from which a stop list was compiled consisting of the lemmas of the 1000 most common words in this corpus. Apart from the three selection requirements, additional manual work consisted of removing, for instance, abbreviations and numerals. The entries in the word list consisted of part-of-speech, inflectional form, meaning, authentic corpus examples and English translation (Carlund et al., 2012).

A disadvantage of the Swedish academic word list is that it contains quite a few words considered to be common in general language. This is probably a consequence of using the vocabulary derived from an easy-to-read corpus as stop list. This is further discussed and accounted for in Section 3.5 of this paper.

[3.4] *The Norwegian academic word list*

The academic word list for Norwegian Bokmål contains 750 words.¹ The list is constructed from the Norwegian Bokmål texts from the DUO Corpus - an academic corpus of more than 100 million words from all the faculties at the University of Oslo: Humanities, Educational Sciences, Medicine, Social Sciences, Mathematics and Natural Sciences, Theology, Law, and Dentistry. The corpus and the method used for generating the list are described in more detail in Johannessen et al. (2016).

The method is a reimplementations of the procedure described by Gardner & Davies (2013), and has four steps where the first one excludes common high frequency words and the other three exclude subject-specific words. The four steps are:

- (i) **Ratio:** To eliminate general high frequency words, a word in the academic DUO corpus must have a higher frequency than in a general corpus. The web corpus NoWaC (Guevara 2010), of 700 million words, was used as a reference corpus of general texts. The most interesting results, after some experimentation, were found by keeping the words that were 2.2 - 2.6 times as frequent in the academic corpus as in the reference corpus.
- (ii) **Range:** A word must occur with a certain frequency in all academic domains. In the DUO Corpus, at least 6 of the 8 faculties should be represented. A range between 30 and 40% was found to give the best results.
- (iii) **Dispersion:** The dispersion measure (see Jullian & Chang-Rodriguez, 1964) is a complementary indication of how "evenly" a word is distributed in a corpus. The measure ranges from 0.01 (the word only occurs in a small part of the corpus) to 1 (even distribution throughout the whole corpus). We found that words in the academic corpus must have a dispersion of at least 0.60.
- (iv) **Discipline Measure:** Like range and dispersion this measure is designed to exclude discipline-specific words. Our calculations showed

[1] Norwegian has two official written varieties: Bokmål and Nynorsk. They are used in different geographical areas and have different origin, as the former is developed from Danish, while the latter was developed as a reaction against the former, and based on dialects. The former is the one used most frequently, and by most people. The national broadcasting cooperation (NRK) is expected to have 25% of their programs in Nynorsk.

that a word should not occur more than 3 – 3.2 times the expected frequency (per million words) in any of the 8 faculties.

As indicated above, different input values resulted in different lists. The choice between the lists was finally decided by testing coverage in two test corpora: one fiction corpus and one academic corpus (Johannessen et al. 2016, Section 5). Two lists resulted from the testing; both with high coverage in the academic test corpus and low coverage in the fiction corpus. As both lists had their advantages, the present Norwegian authors (Johannessen and Hagen) merged the two lists manually together with an experienced lexicographer (Professor Ruth Vatvedt Fjeld at the University of Oslo). The final Norwegian Bokmål Academic Word List has a coverage of 8.1% in the academic corpus and a coverage of 1.3% in the fiction corpus.

[3.5] *Differences in the academic wordlists extracted by different methods*

The Swedish and the Norwegian academic word lists are both extracted from academic corpora using the selection principles described in Sections 3.3 and 3.4. The methods for excluding subject-specific terminology and including words common in several academic subjects are comparable for the two lists. Nevertheless, the process of extraction of the two lists differ both in the method for excluding high-frequency non-academic words and in the size and type of the academic corpora used for extraction.

The Swedish corpus is smaller than the Norwegian one and only covers academic subject-specific texts from the humanities and social sciences, while the Norwegian corpus covers texts from all academic disciplines. Probably because of this difference in subjects, a word such as *korrelasjon* ‘correlation’ in the Norwegian list has no equivalent in the original Swedish list. One might assume that this word is more used in the natural sciences than in the social sciences.

Both corpora consist of doctoral theses and articles, but the Norwegian one also includes master theses. Further, to exclude high-frequency non-academic words, a stop list of high-frequency words extracted from a corpus of easy-to-read texts was used for the Swedish list. The Norwegian method included only words which had a higher frequency in the academic corpus than in a general corpus.

Comparing the two methods and the two lists, we found that they do not in fact differ that much. The Swedish list may still have more high-frequency non-academic words because of the stop list used. On the other hand, the Norwegian academic corpus has more texts from non-academic writers as the corpus consists of theses written by master students not yet properly trained in the aca-

demical tradition.

Coverage tests were carried out for both lists. The Swedish list of 750 words has a coverage of 8.7% in an academic corpus (Carlund et al. 2012) while the Norwegian score derived from measuring coverage in an academic control corpus is 8.1%. Gardner & Davies (2013:19) report that their list yields 13.8% coverage in the COCA academic corpus compared to 7.2% using Coxhead's list on the same corpus. The numbers for the Scandinavian coverage are comparable to those claimed by Nation (2001), who concludes that the vocabulary of academic texts consists of 80% of the most common and frequent words, 8-10% general academic words, 5% subject-specific and technical words and 5% low-frequency words.

A high coverage number in an academic corpus does not, however, necessarily mean that the word list is core academic. A list with many high-frequency non-academic words will obtain higher coverage in an academic corpus than a list with fewer such words. The Gardner & Davies list (2013:19), for instance, also has high coverage scores in the other genres: 8% in a newspaper corpus and 3.4% in a corpus of fiction. In contrast, the Norwegian list has a coverage of only 1.3% in a fiction corpus.

[3.6] *The merging of the Norwegian and the Swedish academic word lists*

The Swedish and Norwegian academic word lists were the basis for making the merged list described in this paper. Our method for merging the lists is described below.

The two monolingual lists were translated in three steps; 1) an automatic translation using the multilingual word list Kelly (Kilgarriff et al. 2014), 2) a manual translation of the words not translated in step 1, 3) a manual check and correction using available monolingual and bilingual electronic dictionaries for the two languages. The two lists were then merged into one, and alphabetically sorted. Each entry was first checked manually once, then double-checked by the present authors using the same dictionaries.

Some entries were unique in one language and non-existent in the target language list. All entries containing a unique entry were controlled for frequency in the target language using a frequency list. If an item was missing from one of the two lists and was identified as one of the 2000 most common words in the source language, the word was excluded from the list unless it was polysemous and had a specific contextual sense. For this purpose we used frequency lists of the 2000 most common words in general speech in Norwegian and Swedish: the NoWaC web corpus for Norwegian and the Blog corpora for Swedish.

We also considered the possibility that unique entries in Norwegian might be the result of a more diverse and balanced corpus subject-wise. In such cases, this was regarded as making a positive contribution to the merged list.

[4] THE MERGED BILINGUAL ACADEMIC WORD LIST

In this section, the characteristics of the bilingual academic word list are described. Comparisons of the relative academic ranking of words between the two languages and various language-specific traits are also dealt with, as are cognates and false friends.

[4.1] *The merged Norwegian-Swedish academic word list*

One interesting point in the merged list is the number of words that differ in academic index rank in the two lists. The merged list consists of 975 word pairs, but only 261 of them are pairs where the two words differ by fewer than 100 rank positions in the respective source language academic word list. Some examples of word pairs that have equivalent ranks are given in Table 1. (The word to the left is always the basis for finding a translation equivalent in the other language, and is marked for language: N=Norwegian, S=Swedish.)

The table shows the word from one list with its rank in the academic frequency list, the equivalent word in the target language and its rank, and finally the English translation.

Source word	Rank	Translation equivalent in target language	Rank	English rendering
aktör (S)	112	aktør (N)	84	'agent'
analys (S)	17	analyse (N)	30	'analysis'
anse (S)	23	anse (N)	111	'consider'
anta (S)	179	anta (N)	137	'assume'
aspekt (S)	80	aspekt (N)	112	'aspect'
avgörande (S)	234	avgjørende (N)	178	'crucial'
orsak (S)	195	årsak (N)	105	'cause', 'reason'
ägna (S)	353	vie (N)	447	'devote'

TABLE 1. Words in Swedish and Norwegian with similar ranks in the list.

We also provide a list of words selected to show how different the two languages can be from each other, see Table 2.

Swedish	Norwegian	English translation
avsedd	tiltenkt	'intended'
beteende	oppførsel	'behaviour'
förefalla	synes	'seem'
ifrågasätta	sette spørsmåltegn ved	'to question'
i synnerhet	særlig	'especially'
jämförelse	sammenlikning	'comparison'
utmaning	utfordring	'challenge'
åsikt	oppfatning	'understanding', 'opinion', 'perception'

TABLE 2. Differences between Swedish and Norwegian academic vocabulary.

[4.2] *Cognates, false friends and remote friends*

When merging two lists from languages that are as close as Swedish and Norwegian, the expectation would be that most words in the academic domain are the same. For words that belong to the common heritage this expectation is perhaps even stronger. And yet, there is a substantial number of words that turn out not to be equivalents in spite of their common origin and their phonological form.

There are many cognates in the lists, though many of them are not equivalent from the point of view of rank in the academic frequency lists. Some examples of cognates are given in Table 3.

Source word	Rank	Translation equivalent in the target language	Rank	English rendering
vardaglig (S)	516	hverdagslig (N)	0	'everyday-like'
variant (S)	624	variant (N)	0	'variant'
variation (S)	333	variasjon (N)	245	'variation'
värde (S)	120	verdi (N)	59	'value'
värdefull (S)	608	verdifull (N)	0	'valuable'
värdera (S)	582	vurdere (N)	0	'evaluate'
värdering (S)	260	vurdering (N)	88	'valuation'

TABLE 3. Some examples of cognates.

The cognates in the two languages have all kinds of origin, from heritage words like *vardaglig/hverdagslig* 'everyday-like' to loanwords from German and derivations from them, like *värde/verdi* 'value', and from Latin, like *variant*. Notice that some of the translations (i.e., the third column) cannot be found in the ac-

ademic list of that language (position 0 in the fourth column). In the cases where Norwegian is the target language, this is probably due to the fact that these are considered as too common to be academic by the Norwegian method, see Section 3.4.

In addition, there are near-cognates, where the two languages have partially similar cognates. Some examples are given in Table 4. We have noticed that in many cases, the academic words in each pair are similar, but not quite the same. They may be compounds or derivational words that have the same second member or suffix, but a different prefix. The second member is often a lexical stem, as in Table 4 a, b, c, e, f, but it also happens that it is a prefix or a first member that they have in common, as in Table 4 d.

	Source word	Rank	Translation equivalent in target language	Rank	English rendering
a	avseende (S)	261	henseende (N)	734	'respect', 'regard'
b	avsikt (S)	474	hensikt (N)	175	'purpose'
c	motsvarande (S)	389	tilsvarende (N)	0	'equivalent'
d	utesluta (S)	590	utelukke (N)	446	'exclude'
e	utlede (N)	661	härleda (S)	0	'deduce'
f	virkemiddel (N)	261	medel	344	'means'

TABLE 4. Some examples of near-cognates.

[4.3] *False friends*

In two languages that have so much common vocabulary historically, but that are separated by national borders, with different public institutions, such as schools and hospitals, different media such as radio, television and newspapers, and different political institutions, it is to be expected that the respective vocabularies will develop in different directions. This is what happens with the phenomenon known as “false friends” (Lamb 1997, Perl & Winter 1972). These are words that look the same (homographs), and probably have the same etymology, but do not mean the same (any more). False friends might also be two or more words from different languages that sound similar, but have different meanings. The words are homographs, homophones or homonyms, but not synonyms (Cambridge Advanced Learner’s Dictionary & Thesaurus).

Table 5 provides some examples of false friends. In the examples a source word is presented with an equivalent in the target language. The English translation is presented (English rendering), followed by a “false friend” in the target. The English translation of the false friend is also presented in the table.

	Source word	Translation equivalent in target language	English rendering	False friend	English rendering
a	begrunne (N)	motivera (S)	'state the reason for'	begrunda (S)	'to ponder'
b	bemärkelse (S)	betydning (N)	'sense'	bemerkning (N)	'remark'
c	erkjenne (N)	inse (S)	'realise'	erkänna (S)	'accept'
d	faglig (N)	fackmässig (S)	'professional, technical'	facklig (S)	'to do with trade union'
e	foregå (N)	försega (S)	'take place, happen'	föregå (S)	'precede'
f	föregå (S)	gå foran (N)	'precede'	foregå (N)	'take place', 'happen'
g	rubrik (S)	overskrift (N)	'headline'	rubrikk (N)	'little section in a newspaper'
h	saknas (S)	mangle (N)	'lack'	savne (N)	'miss', 'long for'
i	tillämpa (S)	anvende (N)	'practice'	tillempa (N)	'adapt'

TABLE 5. Example of "false friends" in Swedish and Norwegian.

[4.4] *Remote friends*

Sometimes cognates exist in the two languages, with almost the same sense, but they are not used in the same way. One of them may be used in an academic way while the other is an everyday word. We present a number of such cases in Table 6. In the examples in the table, a source word is presented with a translation in the target language followed by an English translation. In the last column, a remote friend, a word in the target language that is similar to the source word, is presented. Ranks mentioned in the table are taken from the academic source word lists. Note that no rank has been given for the remote friends, as they are not found in the merged list.

Source word	Rank	Translation equivalent	Rank	English rendering	Remote friend
dock (S)	1	allikevel (N)	233	'though'	dog (N)
ändring (N)	45	förändring (S)	47	'change'	ändring (S)
förändring (S)	47	ändring (N)	45	'change'	forändring (N)
granska (S)	621	undersøke (N)	90	'investigate'	granske (N)
granskning (S)	645	undersøkelse (N)	23	'investigation'	granskning (N)
mån (S)	305	grad (N)	6	'extent'	monn (N)
overordnet (N)	246	övergripande (S)	214	'overriding'	överordnat (S)
samband (S)	51	forbindelse (N)	63	'connection'	samband (S)
sammenfattning (S)	432	oppsamm- mering (N)	302	'summary'	sammenfatning (N)
særegen (N)	511	spesiell/ spesiellt (S)	371/ 575	'characteristic'	säregen (S)
undersøke (N)	90	granska (S)	621	'investigate'	undersöka (S)
undersøkelse (N)	23	granskning (S)	645	'investigation'	undersökning (S)
vie (N)	447	ägna (S)	353	'dedicate'	viga (S)
ägna (S)	353	vie (N)	447	'dedicate'	egne (N)

TABLE 6. Examples of "remote friends" in Swedish and Norwegian.

[4.5] *Phrases*

During the translation process, one of the differences in the source and target languages was that a lexical item in the source language would correspond to a phrase in the target language. Some examples of paraphrasing are given in Table 7.

Source word	Rank	Translation equivalent	English rendering
gjenkjenne (N)	504	känna igen (S)	'to recognize'
ifrågasätta (S)	318	sette spørsmålsteget ved (N)	'to question', 'to dispute'
likhetstrekk (N)	611	gemensamma drag, likhet (S)	'mutual' 'traits'/'similarity'
nedfelle (N)	670	skrive ner (S)	'to enshrine'
omgivande (S)	638	som omgir (N)	'ambient'
omhandle (N)	204	handla om (S)	'to deal'
slutligen (S)	370	til slutt (N)	'finally'
syfta (S)	230	visse (til at)	'to refer to'/'aim to'
oppmerksom (S)	252	gjøre oppmerksom på (N)	'to pay attention to'

TABLE 7. Examples of paraphrasing in the target language.

Since the words in Table 7 lack a one-word translation, it also explains why the word doesn't exist in the target list.

Furthermore, we found a need to distinguish some words in the academic word lists which would typically occur in a phrase but are included in the list as separate words anyway. See examples in Table 8.

Source word	Rank	Translation equivalent	Rank	Phrase	English rendering
anspråk (S)	543	krav (N)	0	göra anspråk på (S)/ gjør krav på (N)	'make a claim'/'claim'
form (S)	10	form (N)	8	i form av/en form av	'in the shape of'/'a form of'
rede (N)	163	redo (S)	0	gjør rede for (N)/ redogöra för (S)	'explain'
säkerhet (S)	629	sikkerhet (N)	0	med (all) säkerhet (S)	'without a doubt'
synnerhet (S)	535	særlig (N)	0	i synnerhet (S)	'especially'
termer (S)	291	når det gjelder (N)	0	i termer av	'in terms of'
utgå (S)	107	gå ut fra (N)	0	utgå från	'assume'

TABLE 8. Examples of words used in phrases.

The fact that these words are part of phrases would argue for an academic word list which would include phrases as well.

[4.6] *Other problems*

There are also other challenges concerning the lack of parallelism between the two languages. For example, it is not always possible to find an appropriate equivalent. An appropriate translation would be one that was equivalent in frequency, context and use. Some words are not commonly used in the target language, are absent or not used in an academic context in the same manner. One such example is the Norwegian *føring* 'constraint', whose nearest translation in Swedish is *restriktion* 'restriction'. No doubt there are contexts where they are equivalent, but while *føring* is a word that can be used in a positive as well as a negative way, the Swedish word has a more negative usage.

Another challenge is the contrast between the origins of the words used. In some cases, one language is more disposed than the other to use loan words. For example, the Norwegian verb *konkludere* 'conclude' has a rarely used Swedish equivalent *konkludera*. In Swedish, *dra en slutsats*, is more commonly used, while the Norwegian noun *konklusjon* 'conclusion' has the rarely used Swedish

equivalent *konklusion*. *Slutsats* is more commonly used in Swedish. Another example is the Norwegian *supplere* ‘supply’, which is a loan word from English, where in Swedish one would use *skaffa/anskaffa* ‘to get’, *tillhandahålla* ‘supply’, *erbjuda* ‘offer’.

We could ask whether the equivalents we have forced into the common Norwegian-Swedish list actually have the same meanings and connotations. This needs to be investigated further.

[5] CONCLUSIONS

As Cobb & Horst (2004) conclude in their paper, “Is there room for an AWL in French?”, in the search for a French academic word list, it seems to be sufficient to master the common vocabulary of French to cover 90% of an academic text. This is an indication that “the acquisition process is able to proceed on a naturalistic basis for learners of French as it is not for learners of English.”

Our belief is that the new merged Norwegian-Swedish academic word list is a valuable resource since neither one of the monolingual word lists is complete. The bilingual list has to some extent remedied this problem since it is enriched by new words. The additional words are either academic from other educational fields or language-specific academic words.

During the work of compiling a merged Norwegian-Swedish academic word list, it has become apparent that academic language use in the two languages differs to some extent. The merged academic word list clearly shows that the two languages do not share exactly the same vocabulary in an academic context and that existing cognates do not necessarily have the same academic status in both languages. Apart from the fact that the word lists are compiled using different methods on differently balanced corpora, some words are only present in one of the source languages since the potentially equivalent words in the target language are not considered to be used in the same way, at least not in an academic context. However, it is just as important for these words to be present in a bilingual list such as this one, since the purpose of the list is to make it easier for language learners or even native speakers of one or the other language to acquire and understand words and phrases in an academic context.

Future work on the word list will include more phrases and further specification of existing lexical entries which form part of phrasal expressions. These examples need to be learned/taught in context.

ACKNOWLEDGEMENTS

We would like to thank two anonymous reviewers for valuable comments and suggestions. We would also like to thank our Swedish, Danish and Norwe-

gian colleagues in the network LUNAS – Language Use in Nordic Academic Settings, and Nordplus language for a lot of inspiration and some funding.

This research has also been partly supported by the Research Council of Norway through its Centre, of Excellence funding scheme, project number 223265.

REFERENCES

- Carlund, Carina, Håkan Jansson, Sofie Johansson Kokkinakis, Julia Prentice & Judy Ribeck. 2012. An academic word list for Swedish - a support for language learners in higher education In *Proceedings of the SLTC 2012 workshop on NLP for CALL*, Lund, 25th October, 2012.
- Cobb, Tom & Marlise Horst. 2004. Is there room for an AWL in French? In B. Laufer & P. Bogaards (eds.) *Vocabulary in a Second Language: Selection, acquisition, and testing*. Amsterdam: John Benjamins, 15-38. http://www.er.uqam.ca/nobel/r21270/cv/awl_F.htm
- Coxhead, Averil. 2000. A New Academic Word List. *TESOL Quarterly*, 34(2), 213-238.
- Gardner, Dee & Mark Davies. 2013. A New Academic Vocabulary List. *Applied Linguistics* 4.
- Guevara, Emiliano Raul. 2010. NoWaC: a large web-based corpus for Norwegian. In *Proceedings of the NAACL HLT 2010 Sixth Web as Corpus Workshop*, Association for Computational Linguistics page 1 - 7.
- Jansson, Håkan, Sofie Johansson Kokkinakis, Judy Ribeck & Emma Sköldberg. 2012. A Swedish Academic Word List: Methods and Data. *Proceedings of the XV Euralex International Congress*, Oslo, August 2012.
- Johannessen, Janne Bondi, Arash Saidi & Kristin Hagen. 2016. Constructing a Norwegian Academic Wordlist. In Calzolari, Nicoletta, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and Stelios Piperidis (eds.). *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Paris: European Language Resources Association (ELRA), 1457-1462.
- Julliard, Alphonse & Eugenio Chang-Rodríguez. 1964. *Frequency Dictionary of Spanish Words*. The Hague: Mouton.
- Kenney, John F. & E. S. Keeping. 1962. *Mathematics of Statistics, Part 1* (3rd ed.). Princeton, NJ: Van Nostrand Reinhold.

- Kilgarriff, Adam, Frieda Charalabopoulou, Maria Gavrilidou, Janne Bondi Johannessen, Saussan Khalil, Sofie Johansson Kokkinakis, Robert Lew, Serge Sharoff, Ravikiran Vadlapudi & Elena Volodina. 2014. Corpus-based vocabulary lists for language learners for nine languages. *Language Resources and Evaluation*, 48:121–163, DOI 10.1007/s10579-013-9251-2.
- Language policy guidelines for the University of Oslo:*
<http://www.uio.no/english/for-employees/support/profile/language/>
- Lamb, Martin. 1997. True Friends, False Friends and Fickle Friends. *Modern English Teacher* 6.2, 29–31.
- Nagy, William. 2007. *Metalinguistic awareness and the vocabulary-comprehension connection*, *Vocabulary Acquisition: Implications for Reading Comprehension*. New The Guilford Press, 52–77.
- Nation, I.S.P. 1990. *Teaching and Learning Vocabulary*, New York: Newbury House.
- Nation, I. S. 2001. *Learning vocabulary in another language*. Ernst Klett Sprachen.
- Nippold Marilyn A. & Lei Sun. 2008. *Knowledge of morphologically complex words: a developmental study of older children and young adolescents*, *Language, Speech, and Hearing Services in Schools*, vol. 39, 365– 73.
- Paquot, Magali. 2010. *Academic Vocabulary in Learner Writing: From Extraction to Analysis*. London & New-York: Continuum, 29–63.
- Perl, Matthias & Roger Winter. 1972. Zum Problem der “false friends”. *Fremdsprachenunterricht* 4, 183–185.
- Ribeck, Judy, Håkan Jansson, & Emma Sköldberg. 2014. Från aspekt till övergripande – en ordlista över svensk akademisk vokabulär. I: *Nordiska studier i lexikografi 12*. Rapport från Konferensen om lexikografi i Norden, Oslo.
- Savický, Petr & Jaroslava Hlaváčová. 2002. Measures of word commonness. *Journal of Quantitative Linguistics*, 9, 215–231.
- Schmitt, Norbert & Cheryl Boyd Zimmerman. 2002. Derivative word forms: what do learners know?, *TESOL Quarterly*, vol. 36, 145–71.
- Scott, Mike & Christopher Tribble. 2006. *Textual Patterns: keyword and corpus analysis in language education*, Amsterdam: John Benjamins, 55.

- Sköldberg, Emma & Sofie Johansson Kokkinakis. 2012. A och O om akademiska ord. Om framtagning av en svensk akademisk ordlista. *Nordiska studier i lexikografi 11. Rapport från Konferensen om lexikografi i Norden, Lund, maj 2011.*
- Spina, Stefania. 2010. *AIWL: una lista di frequenza dell'italiano accademico*. Dipartimento di Scienze del linguaggio, Università per stranieri di Perugia. JADT 2010: 10th International Conference on Statistical Analysis of Textual Data. http://lexicometrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-1317-1326_192-Spina.pdf
- Thorndike, Edward Lee. 1921. *The teacher's word book*. New York, NY, US: Teachers College Press *The teacher's word book*. vi, 134.
- West, Michael Philip. 1953. *A General Service List of English Words*. Longman, London.
- Xue, Guoyi & I. S. P. Nation. 1984. A University Word List. *Language Learning and Communication* 3, 2, 215-229.

URLS

The Academic Keyword List (AKL):

<https://uclouvain.be/en/research-institutes/ilc/cecl/academic-keyword-list.html>

Academic Italian Word List (AIWL):

<http://elearning.unistrapg.it/corpora/AIWL.pdf>

Bokmålsordboka/Nynorskordboka:

<http://ordbok.uib.no/perl/ordbok.cgi?ordbok=begge>

Dicionário de Termos Linguísticos:

<http://www.portaldalinguaportuguesa.org/index.php?action=terminology>

French/English Glossary of Linguistic Terms (Thomas Bearth): http://www-01.sil.org/linguistics/glossary_fe/glossary_index.asp

LäsBarT: <https://spraakbanken.gu.se/resource/lasbart>

Norwegian-Swedish academic word list:

<http://www.tekstlab.uio.no:4000/about>

NoWac Norwegian Web as a Corpus:

<http://www.hf.uio.no/iln/om/organisasjon/tekstlab/prosjekter/nowac/>

Swedish Corpora: <http://spraakbanken.gu.se>

Swedish dictionary SAOB: <http://www.saob.se/so>

Swedish synonym dictionary: <http://synonymer.se>

SwePub: <http://swepub.kb.se/>

Word lists of false friends:

https://sv.wikipedia.org/wiki/Lista_över_falska_vänner_mellan_svenska_och_övriga_nordiska_språk

APPENDIX 1: MONOLINGUAL AND TRANSLATED ACADEMIC WORD LISTS

English

- (i) The Academic Word List (AWL) by Coxhead, see Section 2.1.
- (ii) The Academic Vocabulary List (AVL) by Gardner & Davies, see Section 2.1.
- (iii) The Academic Keyword List (AKL) (Paquot 2010). This list was developed by Magali Paquot within the framework of a research project led by Professor Sylviane Granger at the Centre for English Corpus Linguistics, Université Catholique de Louvain, Belgium. The AKL consists of 930 academic words. The words are reasonably frequent in a wide range of academic texts but relatively uncommon in other kinds of texts. They refer to those activities that characterize academic work, organize scientific discourse and build the rhetoric of academic texts, and thus can be granted the status of academic vocabulary. The word selection is primarily based on keyness.

Portuguese

- (iv) Dicionário de Termos Linguísticos (Maria Francisca Xavier, Maria Helena Mateus, Instituto de Linguística Teórica e Computacional, Associação Portuguesa de Linguística 1990–1992). This is a list on linguistic terminology in Portuguese. The list consists of 14 subsections in various areas within linguistics. Each entry in the list has a definition in Portuguese and information on synonyms as well as translations into English and French.

Italian

- (v) Academic Italian Word List (AIWL) is a frequency list of the most common non-technical words used in written academic communication. The AIWL includes 403 lemmas and 208 of the most common collocations in the written Italian academic lexicon (Spina 2010).

English and ESP

- (vi) French/English Glossary of Linguistic Terms (Thomas Bearth). This glossary contains 7,837 French linguistic terms and 8,059 English linguistic terms. As a glossary, it does not define the terms but simply gives the equivalent(s) in the other language. Lexical and semantic relationships

are displayed for many of the terms in both languages.

APPENDIX 2: CENTRAL TECHNICAL TERMS

Technical term	Explanation	Reference
Raw frequency (RF)	Means absolute frequency, which can be defined as the frequency of an event i being the number n_i of times the event occurred in an experiment or study.	Kenney & Keeping (1962)
Expected and observed frequency	Is a theoretical predicted frequency obtained from an experiment presumed to be true until statistical evidence in the form of a hypothesis test indicates otherwise. An observed frequency, on the other hand, is the actual frequency that is obtained from the experiment.	Wikipedia. "Expected Frequency", http://en.wikipedia.org/wiki/Pearsons_Chi_Square accessed 2017-05-01
Reduced frequency	Average Reduced Frequency (ARF) is a variant on a frequency list that 'discounts' multiple occurrences of a word that occur close to each other, e.g. in the same document.	Savický & Hlaváčová (2002)
Keyness	Describes the quality a word or phrase has of being "key" in its context. Keywords are "items of unusual frequency in comparison with a reference corpus."	Scott & Tribble (2006)
Stop list	A list of words or other data items which, for some special reason, should be ignored or bypassed by a particular data processing operation. If you equip your search engine with a stop list containing a few common words such as "a", "the", and "and", you can decrease the full-text index	Wiktionary. "Stop list", https://en.wiktionary.org/wiki/stop_list , accessed 2017-05-01

size by about 20%.

Dispersion	Measures the dispersion of a word's subfrequencies over n equally-sized subcategories of a corpus.	Julliard & Chang-Rodriguez (1964)
-------------------	--	-----------------------------------

CONTACTS

Sofie Johansson
University of Gothenburg
sofie.johansson@svenska.gu.se

Kristin Hagen
University of Oslo
kristin.hagen@iln.uio.no

Janne Bondi Johannessen
University of Oslo
j.b.johannessen@iln.uio.no